MEAN AND VARIANCE BOUNDS CHARACTERIZATION OF PROBABILITY DISTRIBUTIONS

[1]Naresh Kumar

[1]Research Scholar in Malwanchal University, Indore, Madhya Pradesh, India

[2]Dr. Gaurav Varshney

[2]Research Dean, Statistics, Malwanchal University, Indore, Madhya Pradesh, India

ABSTRACT

We study power-mixture type functional equations in terms of Laplace–Stieltjes transforms of probability distributions on the right half-line [0, ∞). These equations arise when studying distributional equations of the type Z d= X + TZ, where the random variable T ≥ 0 has known distribution, while the distribution of the random variable Z ≥ 0 is a transformation of that of X ≥ 0, and we want to find the distribution of X. We provide necessary and sufficient conditions for such functional equations to have unique solutions. The uniqueness is equivalent to a characterization property of a probability distribution. We present results that are either new or extend and improve previous results about functional equations of compound-exponential and compound-Poisson types. In particular, we give another affirmative answer to a question posed by J. Pitman and M. Yor in 2003. We provide explicit illustrative examples and deal with related topics.

Keywords: distributional equation; power-mixture transform; functional equation; characterization of distributions

INTRODUCTION

We deal with probability distributions on the right half-line [0, ∞) and their characterization properties expressed in the form of distributional equations of the type Z d= X + TZ, where the random variable T ≥ 0 has known distribution, the distribution of the random variable Z ≥ 0 is a transformation of that of X ≥ 0, and we want to find the distribution of X. By using Laplace–Stieltjes (LS) transform of the distributions of the random variables involved, we transfer such a distributional equation to a functional equation of a specific type. Our goal is to provide necessary and sufficient conditions for such a functional equation to have a unique solution. The unique solution is equivalent to a characterization property of a probability distribution.

The impetus for considering ill-specified random variables is that, in most practical situations, no information about the random variable's probability distribution is available. The analyst, lacking adequate information to faithfully model a random variable, often will assume additional knowledge, and hope to safeguard this assumption with sensitivity analysis.

Our proposed definition of ill-specified random variables requires limited amounts of information that often will be available, we will show that our definition also encompasses many of the established procedures for modeling ambiguity.

Characterization results for probability distributions are important part of statistics and probability applications. This includes generic distribution classifiers like characteristic or mean residual life functions as well as specific identification methods like independence of mean and variance estimators for Gaussian distributions; see Galambos and Kotz (2006) and Ahsanullah (2017), and references therein. In this note, we focus on the former and show that the information about doubly quantile censored variance function is sufficient to uniquely determine the distribution up to an additive constant.

The power-mixture functional equations arise when studying power-mixture transforms involving two sii-processes. Here the abbreviation "sii-processes" stands for stationary independent-increments stochastic processes. Think, for example, of the Lévy processes. Consider a continuous time sii-process $(X1(t))t\geq 0$, and let F1,t be the (marginal) distribution of X1(t); we write this as $X1(t) \sim F1,t$.

Moreover, let $X1 := X1(1) \geq 0$ be the generating random variable for the process, so $X1 \sim F1 := F1,1$ uniquely determines the distribution of the process $(X1(t))t\geq 0$ at any time t. Thus, we have the multiplicative semigroup $(\hat{F}\ 1,t(s))t\geq 0$ satisfying the power relation

$$\hat{F}_{1,t}(s) = (\hat{F}_1(s))^t, \quad s,t \geq 0.$$

Here $\hat{F}$ 1,t is the LS transform of the distribution F1,t of X1(t):

$$\hat{F}_{1,t}(s) = \mathrm{E}[e^{-sX_1(t)}] = \int_0^\infty e^{-sx}\, dF_{1,t}(x), \quad s \geq 0$$

LITERATUTE REVIEW

Nagaraja H. (2016), A characterization is a certain distributional or statistical property of a statistic or statistics that uniquely determines the associated stochastic model. This chapter provides a brief survey of the huge literature on this topic. Characterizations based on random (complete or censored) samples from common univariate discrete and continuous distributions, and some multivariate continuous distributions are presented. Characterizations that use the properties of sample moments, order statistics, record statistics, and reliability properties are reviewed. Applications to simulation, stochastic modeling and goodness-of-fit tests are discussed. An introduction to further resources is given.

S. Chukova, B. Dimitrov and Z. Khalil (2012), A concept of the lack-of-memory property at a given time point $c > 0$ is introduced. It is equivalent to the concept of the almost-lack-of-memory (ALM) property of the random variables. A representation theorem is given for the cumulative distribution function of such random variables as well as for corresponding decompositions in terms of independent random variables. It is shown that a periodic failure rate for a random variable is equivalent to the ALM property.

Andrew Weatherbee, Mitsuro Sugita, Kostadinka Bizheva, Ivan Popov, and Alex Vitkin (2016), The distribution of backscattered intensities as described by the probability density function (PDF) of tissue-scattered light contains information that may be useful for tissue assessment and diagnosis, including characterization of its pathology. In this Letter, we examine the PDF description of the light scattering statistics in a well characterized tissue-like particulate medium using optical coherence tomography (OCT). It is shown that for low scatterer density, the governing statistics depart considerably from a Gaussian description and follow the K distribution for both OCT amplitude and intensity. The PDF formalism is shown to be independent of the scatterer flow conditions; this is expected from theory, and suggests robustness and motion independence of the OCT amplitude (and OCT intensity) PDF metrics in the context of potential biomedical applications.

Magdy E. El-Adll (2018), In this paper, characterization of probability distributions by equalities of two different generalized order statistics (gos) or dual generalized order statistics (dgos) is considered. It is proved that, if two different gos or dgos via the same distribution function (df) F are equal, then F has at most two growth points.

MEAN AND VARIANCE BOUNDS

We call a real-valued variable an ill-specified random variable when we do not know the precise probability measure on, but we have enough information to constrain the possible realizations to a finite number of points, sets or bounded intervals, i.e., $x_i \in A_i$, and we can constrain the probability mass assignments on A_i to points or bounded intervals; i.e., $m_i \in M_i, i = 1, \ldots, n$. At least one of the sets or intervals must be nondegenerate in order to have an ill-specified random variable. Furthermore, may overlap. Thus, $P[X = x] \le \sum_{k:x \in A_k} m_k$. If there is a singleton set $A_j = \{x\}$, then there is also a lower bound $m_j \le P[X = x]$. Since in this ill-specified space we cannot determine the precise values of the distribution's mean and variance, we propose procedures for obtaining their optimal bounds.

An unwary approach for obtaining the bounds is to use a limited or no exhaustive-sensitivity analysis. However, this approach does not guarantee optimal bounds. For example, consider the following outcome-probability pairs $\{(11, 0.22), ([11, 14], 0.26), ([12, 15], 0.26), (15, 0.26)\}$, where all 's and 's are crisp except two 's which are only known to be in the given intervals. By stepping through the two intervals in small increments, the mean and variance points in Fig. 1 were generated. Now, as might be suggested by a limited or no exhaustive sensitivity-analysis approach, we calculate the mean and variance points using the high, low, and midpoint values of the two intervals. The results are the nine highlighted points in the figure. Note the minimum variance is not found by the sensitivity-analysis approach. The smallest of the nine variance values found by sensitivity analysis is 2.02, whereas the minimum variance determined with techniques from Section III-B is 1.91.

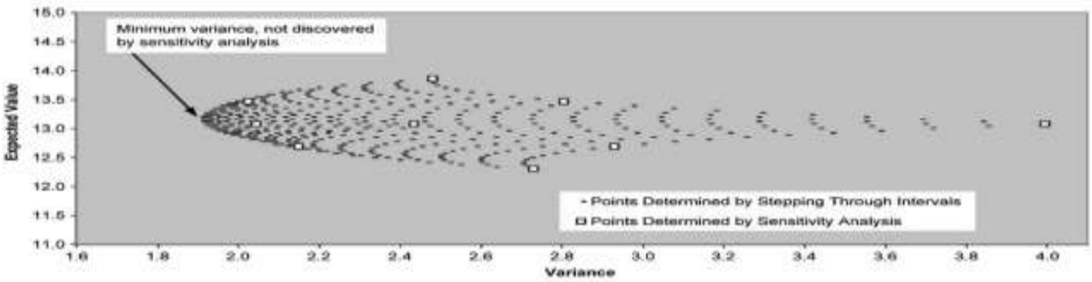

Fig. 1: Typical, irregularly shaped region of E-V points for an ill-specified random variable

The characterization theorems based on conditional second (and higher) moments have been studied in the literature only in specific contexts. For example, in Unnikrishnan Nair and Sudheesh (2010) the authors study how the properties of truncated variance function could result in characterizations for specific classes of non-negative absolutely continuous random variables satisfying certain properties; see Unnikrishnan Nair and Sudheesh (2006) where the required condition, given in Theorem 2.1(iv), is discussed in details. Also, in El-Arishy (2005) the conditional variance characterization in a specific context of some discrete probability distributions is given. Finally, it should be noted that the potential usage of truncated moments as classifiers has been communicated in the literature (e.g., in Laurent, 1974) but we found no direct treatment of this property and the discussion about its potential application.

While the conditional variance function with quantile set trimming seems to be a natural (local) extension of standard variance, it is not considered in the literature as a benchmark framework. This is quite surprising, as the conditional second moments seem to be more natural (e.g. for engineering applications) compared to higher-order moment analysis, e.g. when the tail structure is assessed. In fact, it was shown recently in Jelito and Pitera (2018) that a simple test based on conditional second moments outperforms most of the popular benchmark methods when normality testing is considered. More explicitly, the statistical test power for various choices of popular alternatives (t-student, logistic, and Cauchy distributions) was shown to be bigger compared to reference normality tests based on Jarque–Bera, Anderson– Darling, or Shapiro–Wilk statistics; see Jelito and Pitera (2018,).

The characterization result presented in this note shows that conditional variances might be used for efficient distribution identification and goodness-of-fit testing. In particular, it shows that one could develop efficient statistical testing framework, by controlling the number of included conditional sets with the sample size.

Preliminaries

Let (Ω, Σ, P) be a probability space and let L 0 := L 0 (Ω, Σ, P) denote the set of all (a.s. identified) random variables. For any X ∈ L 0 and A ∈ Σ, such that P[A] ≠ 0, we use

$$\mathrm{Var}[X|A] := \mathbb{E}\left[(X - \mathbb{E}[X|A])^2|A\right]$$

to denote (possibly infinite) conditional variance of X on A; all regularity conditions are taken for granted. For brevity, for any X ∈ L 0 and 0 ≤ a < b ≤ 1 we define a quantile conditioned variance

$$V_X(a, b) := \mathrm{Var}[X|A_X(a, b)]. \quad (2.2)$$

where AX (a, b):= {ω ∈ Ω : X(ω) ∈ [QX (a), QX (b)]} is the quantile set with the lower quantile function QX : [0, 1] → [−∞, +∞] given by

QX (u): = inf {x: u ≤ FX (x)}, u ∈ [0, 1), QX (1): = sup {x: FX (x) < 1}. (2.3)

It should be noted that VX (a, b) is well defined as for 0 ≤ a < b ≤ 1, we get

P [AX (a, b)] = FX (QX (b)) − F l X (QX (a)) ≥ b − a > 0, (2.4)

where F l X (t): = P (X < t), t ∈ R, denotes the Kolmogorov distribution function. Indeed, recalling that QX is the left continuous generalized inverse of the cumulative distribution function, and for any u ∈ [0, 1] we have F l X (QX (u)) ≤ u ≤ FX (QX (u)), we get (2.4). Also, it is worth noting that for u ∈ [0, 1] we get

FX (QX (u)) − F l X (QX (u)) = P [X = QX (u)]. (2.5)

Finally, note that the quantile conditional variance function given in (2.2) is defined up to an additive constant, i.e. for any fixed X ∈ L 0, 0 ≤ a < b ≤ 1, and c ∈ R, we get AX (a, b) = AX+c (a, b), and VX (a, b) < ∞ if additionally 0 < a and b < 1.

MAIN RESULT

In this section we state and prove the main result of this note, i.e., that the information about quantile-based conditional variance is sufficient to characterize the distribution of X up to an additive constant.

Theorem 1.1 Let X, Y be any random variables such that for 0 ≤ a < b ≤ 1 we have VX (a, b) = VY (a, b). Then, there exists c ∈ R such that FX (t) = FY+c (t), t ∈ R, i.e. the laws of X and Y coincide almost surely up to an additive-constant.

Proof. Following the second proof of Theorem 14.1 in Billingsley (2008) we observe that a random variable X has the same distribution as the random variable QX (U), where U(ω):= ω is a uniformly distributed random variable defined on the standard probability space ((0, 1), B((0, 1)), λ) and λ is the Lebesgue measure.

For any $0 \le a < b \le 1$, let $\widetilde{A}(a, b) := \{\omega \in [0, 1] : \omega \in [F_X^l(Q_X(a)), F_X(Q_X(b))]\}$.

Recalling (2.4) and noting that

$$\widetilde{A}(a, b) = \{\omega \in [0, 1] : Q_X(\omega) \in [Q_X(a), Q_X(b)]\} \cup \{F_X^l(Q_X(a))\},$$

we get

$$\mathbb{E}[X|A_X(a,b)] = \mathbb{E}[Q_X(U)|\widetilde{A}(a,b)] = \frac{1}{\mathbb{P}[A_X(a,b)]}\int_{F_X^l(Q_X(a))}^{F_X(Q_X(b))} Q_X(u)du,$$

$$\mathbb{E}[X^2|A_X(a,b)] = \mathbb{E}[Q_X(U)^2|\widetilde{A}(a,b)] = \frac{1}{\mathbb{P}[A_X(a,b)]}\int_{F_X^l(Q_X(a))}^{F_X(Q_X(b))} Q_X(u)^2du.$$

Consequently, using the property

$$|\, V_X(a,b) = \mathbb{E}[X^2|A_X(a,b)] - \mathbb{E}^2[X|A_X(a,b)],$$

We get

$$V_X(a,b) = \frac{1}{2\mathbb{P}[A_X(a,b)]^2} K_X(F_X^l(Q_X(a)), F_X(Q_X(b))),$$

where $K_X : [0,1]^2 \to [-\infty, +\infty]$ is an exchangeable function ($K(x,y) = K(y,x)$) given for $x \le y$ by

$$\begin{aligned} K_X(x,y) &:= 2(y-x)\int_x^y Q_X^2(u)du - 2\left(\int_x^y Q_X(u)du\right)^2 \\ &= \int_x^y\int_x^y \left(Q_X^2(u) + Q_X^2(v)\right) dudv - \int_x^y\int_x^y 2Q_X(u)Q_X(v)dudv \\ &= \int_x^y\int_x^y \left(Q_X(u) - Q_X(v)\right)^2 dudv. \end{aligned}$$

Now, let ΔK_X denote the mixed second difference of K_X,

$$\Delta K_X(x_1,x_2,y_1,y_2) := K_X(x_1,y_1) + K_X(x_2,y_2) - K_X(x_1,y_2) - K_X(x_2,y_1).$$

For $0 \le x_1 < x_2 < y_2 < y_1 \le 1$ we obtain

$$\Delta K_X(x_1,x_2,y_1,y_2) = 2\int_{x_1}^{x_2}\int_{y_2}^{y_1} \left(Q_X(u) - Q_X(v)\right)^2 dudv.$$

Now, let ΔKX denote the mixed second difference of KX,

$$\Delta K_X(x_1,x_2,y_1,y_2) := K_X(x_1,y_1) + K_X(x_2,y_2) - K_X(x_1,y_2) - K_X(x_2,y_1).$$

For $0 \le x_1 < x_2 < y_2 < y_1 \le 1$ we obtain

$$\Delta K_X(x_1,x_2,y_1,y_2) = 2\int_{x_1}^{x_2}\int_{y_2}^{y_1} \left(Q_X(u) - Q_X(v)\right)^2 dudv.$$

Now, we show that the intervals where the quantile function Q is constant are determined by the conditional variance V. Let $\phi, \psi : [0,1] \longrightarrow [0,1]$ be given by $\phi(0) = 0$, $\psi(1) = 1$ and

$$\varphi(u) := \sup\{x : 0 \le x < u, V_X(x,u) > 0\}, \quad \psi(u) := \inf\{x : 1 \ge x > u, V_X(u,x) > 0\}.$$

For $u \in [0,1]$ such that $\phi(u) < \psi(u)$, the quantile function QX is constant on $(\phi(u), \psi(u)]$, and for any $v \in (0,1)$ such that $v \ne u$ and QX (v) = QX (u), we get $v \in [\phi(u), \psi(u)]$. Therefore, for $u \in (0,1)$, we get $\phi(u) = F^l_X(Q_X(u))$ and $\psi(u) = F_X(Q_X(u))$. Consequently, recalling (3.1), for $0 < a < b < 1$, we get

$$K_X(\varphi(a), \psi(b)) = (\psi(b) - \varphi(a))^2 V_X(a,b).$$

Furthermore, for ϕ(u) > 0 and small ε > 0, we get QX (ϕ(u) − ϵ) < QX (u), which implies

$$\varphi(u) = F^{l}(Q_X(u)) \geq F(Q_X(\varphi(u) - \epsilon)) \geq \varphi(u) - \epsilon.$$

Since F (QX (ϕ(u) − ϵ)) = ψ(ϕ(u) − ϵ), we get limv→ϕ(u)− ψ(v) = ϕ(u). Next, since KX (a, b) is continuous, for b such that ϕ(b) > 0, we get KX (a, ϕ(b)) = limv→ϕ(b)− KX (a, ψ(v)) and, due to (3.5), for ϕ(a) < ϕ(b), we obtain

$$K_X(\psi(a), \psi(b)) = \lim_{v \to \psi(a)^+} (\psi(b) - \varphi(v))^2 V_X(v, b).$$

Note that if ψ(a) = ψ(b), then K(ψ(a), ψ(b)) = 0, and if ϕ(b) = ϕ(a), then K(ϕ(a), ϕ(b)) = 0.

Now, if ϕ(u) < ψ(u), then the quantile function QX (·) is constant on (ϕ(u), ψ(u)), so both KX (·, ψ(u)) and KX (ϕ(u), ·) are linear on (ϕ(u), ψ(u)). Combining all facts, we get that integrals KX and ∆KX are determined by the conditional variance VX.

Now, let us assume that Y is such that VX (a, b) = VY (a, b) for all 0 ≤ a < b ≤ 1. Since QX and QY are left-continuous and nondecreasing, the equality ∆KX = ∆KY implies that QX (b)−QX (a) = QY (b)−QY (a) for all a, b ∈ [0, 1]. This concludes the proof.

Theorem 1.1 could be easily extended to the multivariate case e.g., by using information about conditional variances for all linear combination of marginal random variables. In the following theorem we use ⟨·, ·⟩ to denote the standard Euclidean inner product operator.

Theorem 1.2 Let X, Y be any n-dimensional random vectors such that for 0 ≤ a < b ≤ 1 and α ∈ R n we have V⟨α,X⟩(a, b) = V⟨α,Y⟩(a, b). Then, there exists c ∈ R n such that FX (t) = FY+c (t), t ∈ R n , i.e. the laws of X and Y coincide almost surely up to an additive shift.

The proof of Theorem 1.2 follows directly from Theorem 1.1 combined with Theorem 19 from Galambos (1995). To conclude, let us present two simple remarks which outline potential application of Theorem 1.1; similar remarks are true for the multivariate case.

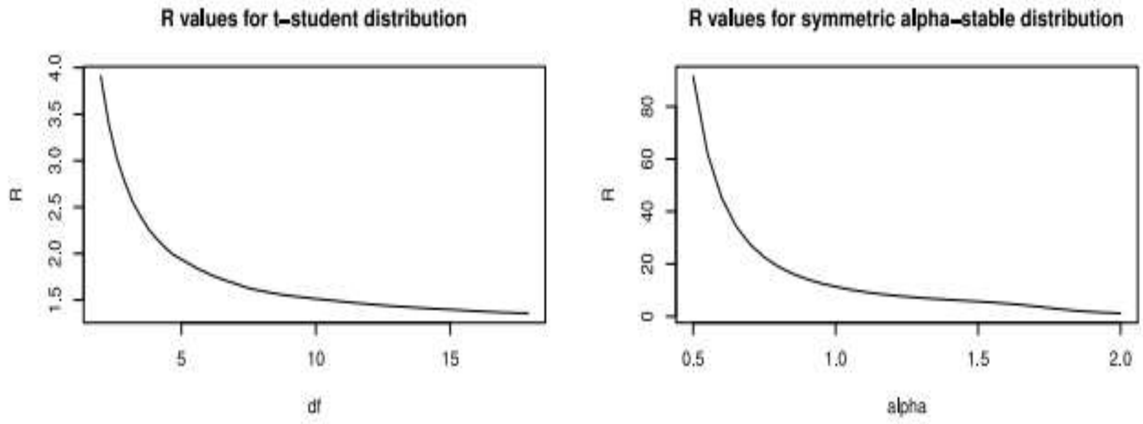

Fig. 2: The plot illustrates the values of R = VX (0.1, 0.3)/VX (0.3, 0.7) under the assumption that X has t-student (left) or symmetric α-stable (right) distribution. R is presented as a function of the underlying parameters: df for t-student (left) and α for symmetric α-stable (right). The values were obtained using Monte Carlo samples of size 10 000 000. In both cases R is a decreasing function of the underlying parameter.

Remark 1.1 (Statistical Goodness-of-fit Testing). As quantile-based conditional variances are easy to estimate and could be used to uniquely classify the distribution (up to an additive constant), they are a natural candidate for goodness-of-fit (shape) statistical testing. In practical applications, it is reasonable to choose a fixed set of specific quantile conditioned sets and then compare conditional variances with the theoretical variances coming from the reference distribution. By introducing various quantile splits and appropriate ratios one might check certain distributional properties rather than the full fit. For example, the comparison of Vˆ X (a, b) and Vˆ X (1 − a, 1 − b) for any $0 \leq a < b \leq 1$ might be used to test distribution symmetry. Also, for a < 0.5, the tail set conditional variances Vˆ X (0, a) and Vˆ X (1 − a, 1) might be compared with the central set conditional variance Vˆ X (a, 1 − a) in order to assess heaviness of the distribution tail.

In fact, exemplary normality testing framework based on conditional variance estimation has been recently introduced in Jelito and Pitera (2018). Using the fact that VX (0, 0.2) = VX (0.2, 0.8) = VX (0.8, 1) for Gaussian random variables we can define the test statistic

$$N := \frac{\sqrt{n}}{1.8} \cdot \frac{\hat{V}_X(0, 0.2) + \hat{V}_X(0.8, 1) - 2\hat{V}_X(0.2, 0.8)}{\hat{V}_X(0, 1)},$$

where Vˆ X (a, b) refers to sample conditional variance constructed by sorting the sample, taking appropriate subset of observation, and applying standard sample variance estimator.1 In Jelito and Pitera (2018), it is shown that the power of related normality test for various choices of popular symmetric alternatives (e.g. t-student, logistic, and Cauchy distributions) is surprisingly big. In particular, test statistic N outperforms popular alternatives like Jarque–Bera, Anderson–Darling, or Shapiro–Wilk tests for samples of size 20, 50, 100, and 250; see Jelito and Pitera, 2018, Also, it is easy to show that N is asymptotically normal.

CONCLUSION

Conditional variances could be also used for parameter fitting. While being relatively simple to establish, the framework based on conditional second moments is much more flexible compared e.g. to method of moments. This is due to the fact that one could consider multiple choices of quantile intervals (a, b) and take their linear combinations; note that sample quantile conditional variance estimators are consistent. To illustrate this, let us consider the ratio R := VX (0.1, 0.3)/VX (0.3, 0.7) for two distribution families: t-student and symmetric α-stable; see Ahsanullah (2017) for details. In Fig. 2, we present the values of R as a function of degrees of freedom (df) and stability index (α) parameters, respectively; note that R is invariant to affine transformations of X. One could see that in both cases R is monotone wrt. parameter change, so that is could be used for parameter identification.

REFERENCES

1. Nagaraja H. (2016), Characterizations of Probability Distributions. In: Pham H. (eds) Springer Handbook of Engineering Statistics. Springer Handbooks. Springer, London. https://doi.org/10.1007/978-1-84628-288-1_4

2. S. Chukova, B. Dimitrov and Z. Khalil (2012), A Characterization of Probability Distributions Similar to the Exponential, Vol. 21, No. 3 (Sep., 1993), pp. 269-276 (8 pages)

3. Magdy E. El-Adll (2018), CHARACTERIZATION OF DISTRIBUTIONS BY EQUALITIES OF TWO GENERALIZED OR DUAL GENERALIZED ORDER STATISTICS, Volume (26) - Issue (3) - 2018

4. Andrew Weatherbee, Mitsuro Sugita, Kostadinka Bizheva, Ivan Popov, and Alex Vitkin (2016), Probability density function formalism for optical coherence tomography signal analysis: a controlled phantom study, Volume 14.

5. Abramowitz, M., and Stegun, I. A. (1970). Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables, Dover, New York, USA.

6. Ahsanullah, M. (2017). Characterizations of Univariate Continuous Distributions, Atlantis Press, Paris, France. DOI: https://doi.org/10.2991/978-94-6239-139-0

7. Ahsanullah, M., and Shakil, M. (2012). A note on the characterizations of Pareto distribution by upper record values, Communications of the Korean Mathematical Society, 27 (4), 835 - 842. DOI: https://doi.org/10.4134/CKMS.2012.27.4.835

8. Ahsanullah, M., Nevzorov,V. B., and Shakil, M. (2013). An Introduction to Order Statistics, Atlantis Press, Paris, France. DOI: https://doi.org/10.2991/978-94-91216-83-1

9. Ahsanullah, M., Kibria, B. M. G., and Shakil, M. (2014). Normal and Student´s t Distributions and Their Applications, Atlantis Press, Paris, France. DOI: https://doi.org/10.2991/978-94-6239-061-4

10. Shakil, M., Ahsanullah, M., and Kibria, B. G. (2018). On the Characterizations of Chen's Two-Parameter Exponential Power Life-Testing Distribution, Journal of Statistical Theory and Applications, 17 (3), 393 - 407. DOI: https://doi.org/10.2991/jsta.2018.17.3.1

11. Nevzorov, V. B. (2001). Records: Mathematical Theory, Translation of Mathematical Monograph, American Mathematical Society, Rhode Island, USA. DOI: https://doi.org/10.1090/mmono/194

12. Nagaraja, H. (2006). Characterizations of Probability Distributions, In Springer Handbook of Engineering Statistics (pp. 79 - 95), Springer, London, UK.

13. Koudou, A. E., and Ley, C. (2014). Characterizations of GIG laws: A survey, Probability surveys, Volume 11, pp. 161 - 176. DOI: 10.1214/13-PS227. DOI: https://doi.org/10.1214/13-PS227

14. Galambos, J., and Kotz, S. (1978). Characterizations of probability distributions. A unified approach with an emphasis on exponential and related models, Lecture Notes in Mathematics, 675, Springer, Berlin, Germany. DOI: https://doi.org/10.1007/BFb0069530